

Zugang zu Forschungsdaten

Die Notwendigkeit von Richtlinien für Data Sharing bei Fachzeitschriften Das JoRD-Projekt der Universität Nottingham

Paul Sturges, Marianne Bamkin, Jane H. S. Anders, Azhar Hussain

Einführung

Allgemein herrscht Einigkeit darüber, dass das Teilen von Forschungsdaten sowohl für Transparenz in der Forschung als auch für mögliche Folgenutzung von Bedeutung ist. Die überwiegend positiven Kommentare zugunsten des Sharing können jedoch darüber hinwegtäuschen, dass die Mechanismen zu einer effektiven Umsetzung noch immer Gegenstand von Debatten und keineswegs ein funktionierender Aspekt der Forschung sind. Das erste große Problem im Zusammenhang mit dem Teilen besteht darin, dass mindestens zwei der drei Interessengruppen es zwar mit dem Kopf, nicht aber mit dem Herzen unterstützen. Die Forscher selbst neigen dazu, an „ihren“ Daten zu kleben, was ihnen auch kein noch so rationales Argument für das Teilen austreiben kann. Die Position der Herausgeber erklärt sich aus der Tatsache, dass ihre Einnahmen direkt an Erhalt und Sicherung des Urheberrechts für wissenschaftliche Artikel gebunden sind, die sich durchaus auch auf das profitable „Sharing“ der entsprechenden Forschungsdaten erstrecken können. Das JoRD-Projekt an der Universität Nottingham, über das wir hier berichten, bestätigte die Skepsis, in welchem Ausmaß das Prinzip des Data Sharing in der Praxis angewendet wird. Die Untersuchung konzentrierte sich darauf, wie wissenschaftliche Fachzeitschriften das Teilen verlangen und zulassen. Wir untersuchten die Richtlinien von Zeitschriften zum Data Sharing in der Erwartung, diese würden unter anderem eine gute Orientierung über Datenstrukturen und Metadaten liefern und die Autoren auf geeignete, Internet-verbundene Speicher führen. Tatsächlich fanden wir, dass die gegenwärtigen Data-Sharing-Richtlinien bei Zeitschriften, höflich formuliert, uneinheitlich und inkonsequent waren. Es folgte eine qualitative Untersuchung, um die Ansichten der Interessengruppen einzuschätzen; diese bestätigten den Eindruck, bot aber auch Perspektiven, wie sich effektives Data Sharing fördern ließe. Im Laufe der Untersuchung ergab sich ein zweites Problem, das die Einführung des Teilens verkom-

In der größeren Forschungs-Community, auch in Bibliotheken und anderen Informations-Verwahrungsorten, herrscht zunehmend Einigkeit über die enorme Bedeutung des Data Sharing, sowohl im Hinblick auf Transparenz als auch auf mögliche Folgenutzung. Logischerweise sollte das Sharing in Form von Daten geschehen, die in geeigneten Speichern hinterlegt sind, die wiederum mit geeigneten Zugangsinstrumenten wie beispielsweise Bibliothekskatalogen verlinkt sind. Dabei kommt den Fachzeitschriften, in denen die wissenschaftlichen Arbeiten publiziert werden, eine maßgebliche Rolle zu. Das JoRD-Projekt an der Universität Nottingham untersuchte den derzeitigen Stand der Richtlinien über Data Sharing durch eine statistische Erhebung stichprobenartig ausgewählter Titel. Abgefragt wurden Ansichten und Praktiken von Interessengruppen, darunter die Wissenschaftler und ihre Geldgeber, Herausgeber und Redakteure. Das Projekt zeigte, dass einige der befragten Zeitschriften durchaus über Richtlinien für Data Sharing verfügten; doch zum einen sind diese Zeitschriften in der Minderheit, zum zweiten befürworteten die Richtlinien zwar das Sharing, ohne es jedoch ausdrücklich zu verlangen. Viele der untersuchten Richtlinien besagten kaum etwas über standardisierte Formate für Daten und Metadaten oder die Nutzung von Datenspeichern. Wenn es echtes Data Sharing geben soll, sind Initiativen gefragt, welche die Zeitschriften dazu ermuntern, Richtlinien für das Sharing in genau spezifizierter und angemessener Form einzuführen.

Schlagerworte: Forschungsdaten, Data Sharing, Zeitschriften, Richtlinien

There is a growing consensus in the broader research community, including libraries and other information repositories, that sharing of research data is vital both for transparency and possible reuse. Logically the sharing should be in the form of data held in suitable repositories which is linked to effective access points such as library catalogues. The journals in which the research appears have a central role in this process. The JoRD Project at Nottingham University investigated the current state of journal data sharing policies through a survey of sample titles, and explored the views and practices of stakeholders including the research community and its funders, publishers and editors. The project identified that although a percentage of journals did have a policy on data sharing, they were in a minority, and policies generally encouraged good practice rather than made it a firm requirement. Many of the policies examined had little to say on standardised formats for data, metadata, or the use of data repositories. If there is to be genuine data sharing, initiatives to encourage journals to set out policies that mandate sharing in well-specified and appropriate forms are essential.

Keywords: research data; data sharing; journals; policies.

pliziert, nämlich die Frage, wie Daten in diesem Zusammenhang definiert werden. Die Diskussion über Daten betrifft ein weit komplizierteres Feld von Ressourcen als auf den ersten Blick ersichtlich. Unser Ausgangspunkt war eine Definition der Royal Society (2012) über Daten als: „qualitative oder quantitative Aussagen oder Zahlen, die sachlich sind (oder als solches angenommen werden). Diese können Roh- oder Primärdaten (z.B. aus direkten Messungen) oder von Primärdaten abgeleitet sein, sind aber, über die Berechnung hinaus, noch kein Produkt von Analyse oder Interpretation.“ Tatsächlich fanden wir, dass sich tendenziell alles, was über Data Sharing diskutiert oder in Erörterungen aufgeführt wurde, auf Software, Video, Geodaten, geologische Landkarten, Ontologien, Web-Inhalte, Datenmodelle und sehr viel mehr bezog. Darüber hinaus erwies es sich als unmöglich, die sogenannten ergänzenden Materialien, die zusammen mit den Daten hinterlegt werden könnten, völlig zu ignorieren. Zum Umgang mit ergänzenden Materialien hat die NISO/NFAIS (NISO, 2013) kürzlich Empfehlungen herausgegeben, die dem Problem mangelnder Anleitung bezüglich Auswahl, Lieferung, Findungshilfen und geplanter Konservierung begegnen sollen und nebenbei auch den Umfang möglicherweise einbezogener Materialien deutlicher machen. Ohne solche Materialien zu ignorieren, haben wir versucht, uns auf die wesentlichen Daten zu konzentrieren, welche die Resultate erhärten.

Übersicht über die Literatur

Es gibt viele maßgebliche Erklärungen zugunsten von Data Sharing. Der International Council for Science (ICSU, 2004), die Organization for Economic Cooperation and Development (OECD, 2007) und die britische Royal Society (Royal Society, 2012) haben sich eindeutig zu dem Thema geäußert; sie fordern Offenheit und freien Zugang zu Daten aus öffentlich geförderter Forschung. In ähnlicher Weise erwarten auch Körperschaften, die Finanzmittel gewähren, wie die US National Academy of Sciences (2003), Daten offen zugänglich zu machen. Die Brussels Declaration (STM, 2007) lässt aber auch das Unbehagen des Verlagswesens über das offene Hinterlegen akzeptierter Manuskripte in rechtlich geschützten Archiven erkennen. Schon frühere Untersuchungen beschäftigten sich mit möglichen Richtlinien über Data Sharing für wissenschaftliche Fachzeitschriften. Mitte der 1990er Jahre befragte McCain (1995) 850 Fachzeitschriften, nur bei 132 bestanden identifizierbare Richtlinien. Eine kleinere Umfrage unter medizinischen Fachzeitschriften durch Shriger et al. (2006) ergab widersprüchliche Herangehensweisen und wenig klare

Anleitung. Seither wurden verschiedene wichtige wissenschaftliche Abhandlungen veröffentlicht durch Piwowar, zumeist mit Chapman (darunter Piwowar und Chapman, 2008b, Piwowar, 2010; Piwowar und Chapman 2010a; Piwowar und Chapman 2010b). Die vielleicht wichtigste ist Piwowar und Chapman 2008a, die auf McCains eingehender Untersuchung der Richtlinien anhand der Daten über Genexpression-Mikroarrays aufbaut. Der Artikel klassifiziert die Richtlinien nach ihrer Stärke (stark, schwach, nicht-existent); der Beziehung zwischen der Stärke der Richtlinien zum Rating des Einflusses der Zeitschrift und der Zahl der identifizierbaren Fälle, in denen Daten eingereicht wurden. Später erbrachte das PARSE-Projekt (Kuipers und van der Hoeven, 2009) hilfreiche Daten über die Haltung zum Data Sharing, ebenso einen klaren Standpunkt zu erforderlichen Maßnahmen (Smit, 2011; Smit und Gruttemeier, 2011). Und es gibt die Arbeit von Stodden et al. (2013), die auf einer sehr ähnlichen Untersuchung wie der unseren beruht und mehr oder weniger gleichzeitig erstellt wurde, sich aber auf das Code Sharing konzentriert, durch das computerbezogene Ergebnisse repliziert werden können.

Methoden

Auf den Websites einer repräsentativen Auswahl von fast 400 wissenschaftlichen Fachzeitschriften wurde nach Daten-Richtlinien gesucht. Wurden solche Richtlinien lokalisiert, unterteilten wir sie in Kategorien wie beispielsweise: Welche Daten werden wann und wo hinterlegt; der Zugang zu Daten; die Art von Daten; die Kontrolle von Data Compliance, die Folgen einer Non-Compliance und die Stärke der Richtlinien auf der Grundlage der Definition stark und schwach. Diese wurden anschließend zum Vergleich in eine Matrix eingegeben. Wenn sich auf der Website einer Zeitschrift keine Richtlinien fanden, wurde dies auf der Matrix verzeichnet. Im ersten Schritt der Analyse untersuchten wir einige individuelle Richtlinien im Detail und fügten so lange auf diese Weise untersuchte weitere Richtlinien hinzu, bis wir keine neuen Features mehr entdeckten. Dieses Vorgehen lieferte einen Satz von Kriterien, die für die Analyse aller verbleibenden Richtlinien verwendet werden. Unsere Resultate basieren auf der Anwendung dieser Kriterien. Ein anderer Blickwinkel wurde durch das Einbeziehen von Interessengruppen gewährleistet, welches die Lektionen aus der Prüfung der Zeitschriften ergänzte, dabei wurden einfache qualitative Methoden angewendet, um die Sichtweise wichtiger Akteure zu etablieren. Zu Prinzipien des Data Sharing befragt wurden Verleger, Forschungsförderer, Datendienste, Biblio-

thekare, Forschungsleiter und Manager. Dies erfolgte in mehreren Schritten, unter Verwendung der gegenstandsbezogenen Theoriebildung. Zunächst wurden die Ansichten von Personen aus dem britischen Verlagswesen zu den Prinzipien des Data Sharing, der Motivation für Veränderung und die Herausforderung, die diese Veränderung bedeutet, eingeholt. Die einzelnen Befragten wurden bewusst ihrer Expertise wegen ausgewählt. Sie kamen aus unterschiedlichen Bereichen des Verlagswesens, von großen bis zu klei-

JASA June	Computational Articles	Code Publicly Available
1996	9 of 20	0%
2006	33 of 35	9%
2009	32 of 32	16%
2011	29 of 29	21%

doi:10.1371/journal.pone.0067111.t001

Die Arbeit von Stodden et al. (2013), konzentriert sich auf das Code Sharing, durch das computerbezogene Ergebnisse repliziert werden können.

nen Häusern, von Abonnements-Publikationen bis zu Open-Access-Unternehmen, dazu kamen Vertreter von Forschungsförderern und Datendiensten, aber auch Forschungs-Administratoren und Manager. Im späteren Verlauf des Projekts führten wir noch Interviews mit vier Vertretern akademischer Bibliotheken. Da die aus den Interviews gesammelten Daten einseitig die Sichtweise von Zeitschriften-Redakteuren und -Herausgebern widerspiegelten, musste auch die Meinung von Forschern und Autoren einbezogen werden. Wir organisierten eine Fokusgruppe aus britischen Forschern. Die Teilnehmer wurden nach dem Schneeballverfahren ausgewählt, anfänglich durch einen Kontakt aus einem wissenschaftlichen Debattenforum. Sie repräsentierten einen unterschiedlichen geistes- und naturwissenschaftlichen Hintergrund. Die Resultate aus den Diskussionen der Fokusgruppe und Hinweise aus dem Überblick der Literatur nutzten wir anschließend, um Fragen für eine offene Umfrage unter Forschern zu formulieren, die einem Monat lang im Rahmen des Projekt-Blogs online gestellt wurden (willkürliche Auswahl). Wir erhielten Antworten von 70 Forschern aller akademischen Bereiche und aus allen Erdteilen; ihr Tätigkeitsgebiet erstreckte sich auf insgesamt 36 verschiedene Wissenschaftsbereiche. Nach jeder Stufe der Datensammlung wurden die Daten offen codiert und Antwortmuster identifiziert; die entstehenden Kategorien ermöglichten den Vergleich der Sichtweise über das gesamte Spektrum aller Beteiligten.

Ergebnisse

Die Umfrage bei Zeitschriften

Die derzeitige Handhabung des Data Sharing bei einzelnen Zeitschriften erwies sich, höflich formuliert, als uneinheitlich und inkonsequent. In einem Umfeld, in dem Aussagen und Richtlinien für das Teilen von Daten sprechen, wäre es nicht ungerecht, die derzeitige Lage als völlig unangemessen zu beschreiben. Ungefähr die Hälfte der befragten Zeitschriften hatte überhaupt keine Richtlinien für Data Sharing. Von den vorgefundenen Richtlinien waren mehr als drei Viertel nach der Definition von Piwowar und Chapman schwach, das verbleibende Viertel stark (76%, 24%). Bedeutsamer war, dass die Zeitschriften mit dem höchsten Impact-Faktor auch die stärksten Richtlinien zeigten. Nicht nur bestanden bei weniger Zeitschriften mit geringerem Impact überhaupt Richtlinien zum Data Sharing, sondern dieses Data Sharing war auch seltener obligatorisch. Generell hieß es dort nur, Autoren wären ja vielleicht interessiert, ihre Daten zu teilen. Wir fragten nach identifizierten Richtlinien, um herauszufinden, ob sie eine Klausel dazu enthielten, welche Daten möglicherweise mit einem Artikel verlinkt waren, wo die Daten hinterlegt und wann sie im Prozess der Veröffentlichung zur Verfügung gestellt werden sollten.

Einige Richtlinien spezifizierten nicht, welche Arten von Daten hinterlegt werden sollten. So waren Datensätze, Multimedia, Einzelexemplare, Muster oder Material die am häufigsten genannten Arten von Daten. Auch Strukturen, Protein oder DNA-Sequenzierung und Programmcode oder Software wurden genannt, aber deutlich seltener. Viele Richtlinien waren nicht spezifisch in der Verwendung der Begriffe, der unterstützenden Information oder unspezifischer und sonstiger Daten. Andere Richtlinien unterschieden zwischen Daten, die integraler Bestandteil des Artikels waren, und ergänzenden Daten. Ergänzende Daten könnten die Qualität des Artikels erhöhen, waren aber zur Untermauerung des darin vorgestellten Arguments nicht entscheidend wichtig. Tatsächlich verlangten sieben Prozent der Richtlinien, den Umfang der ergänzenden Daten zu beschränken oder sie erst nach der Diskussion einzufügen.

Noch wichtiger ist, dass nur wenige Richtlinien genau besagten, wo die Daten hinterlegt werden sollten. Einige wenige sprachen über eine Ablage, blieben aber vage in der Frage, wo dies geschehen sollte (7%). Andere (17%) beriefen sich auf die Nutzung eines Speichers, nannten aber keinen spezifischen Speicher. Nur 15% nannten einen spezifischen Speicher. Aussagen über den erwarteten Zugang beispielsweise fehlten; nur 28 der 371 abgefragten Richtlinien besagten



Das Trentgebäude im University Park Campus Nottingham. (aus Wikipedia)

etwas darüber. Davon war bei vier von freiem Zugang die Rede, bei zweien von Open Access und bei 18 von kostengünstigem Zugang. Das Bedenklichste war vielleicht, dass nur eine Richtlinie verlangte, auch Metadaten zu hinterlegen. Auch in der Frage, wann die Daten hinterlegt werden sollten (vor der Veröffentlichung oder zum Zeitpunkt der Publikation), gab es weder Einheitlichkeit noch Richtung. Laut 51% der Richtlinien, die spezifisch etwas darüber besagten, sollten die Daten mit der Einreichung des Artikels zur Verfügung stehen, weitere 23% verlangten die Daten für den Peer Review. Laut den restlichen 23% der Richtlinien reichte es aus, die Daten zu einem späteren Zeitpunkt zu hinterlegen, in der Regel zum Zeitpunkt der Veröffentlichung. Insgesamt fanden wir nur wenige Richtlinien (bei lediglich der Hälfte der befragten Zeitschriften), davon wiederum die Hälfte schwach und unübersichtlich. Die Schwäche lässt sich daran belegen, dass nur bei 10% von Strafen bei Nichteinhalten der Richtlinien die Rede war.

Konsultation der Akteure

Unter den verschiedenen befragten Akteuren, Fokusgruppen und Online-Befragten herrschte nur geringes Verständnis füreinander. Die Akteure machten Aussagen über das Vorgehen und die Ansichten der jeweils anderen und hatten sich offensichtlich wenig bemüht, das allgemeine Umfeld zu untersuchen. Auch wenn die Akteure bemüht waren, den Eindruck zu erwecken, als befürworteten sie das Data Sharing, und dessen Vorteile bereitwillig anerkannten, brachten sie alle Einwände und Bedenken zum Ausdruck

und nannten Hindernisse für das Teilen der Daten. Beispielsweise wurde aus den Kommentaren der Forscher in der Fokusgruppe und bei der Online-Befragung deutlich, dass sich alle in der Erwartung einig waren, dass die Daten geteilt würden. Gleichzeitig zeigte die Online-Umfrage eine weniger positive Wirklichkeit. Rund 40% der Befragten gaben an, anderen keinen Zugriff auf die Daten zu gewähren, die übrigen teilten sie zumeist ausschließlich mit Mitarbeitern und Kollegen. Vom Instinkt her teilen Forscher noch nicht: dies unterstreicht die Notwendigkeit klarer Richtlinien, wenn sich Verhalten, Aufmerksamkeit und Einhalten der Richtlinien von Forschungsförderern und Herausgebern ändern sollen. Wie bereits erwähnt, sollten die Richtlinien in dieser Frage eindeutig sei, wenn sie wirklich etwas bewirken sollen. Die Herausgeber, die aufgerufen sind, ihren Autoren auf ihren Websites und auf den Seiten ihrer Zeitschriften die Richtlinien darzulegen, zeigen sich tatsächlich unsicher angesichts der Möglichkeiten der heutigen digitalen Infrastruktur, Daten zuverlässig mit Artikeln zu verlinken, wenn die Daten auf verschiedene Datenbanken und andere Ablageorte verteilt sind. Andere waren nicht sicher, ob ihre eigenen Datenbanken angesichts des wachsenden Umfangs von Forschungsdaten, die größere Speicherkapazität verlangen, als alternative Speicherorte ausreichen. Das impliziert, dass Forschungseinrichtungen und Forschungsförderer die Chance haben, die Frage der Archivierung in die Hand zu nehmen, und zwar notwendigerweise mit klaren, durchsetzbaren Richtlinien und klaren, benutzerfreundlichen Speicherorten und Verfahren.

Die Konsultation brachte weitere Problemstellungen ans Licht. Sowohl Forscher als auch Herausgeber betrachteten es als schwierig, Daten im dem ursprünglichen Zustand ihrer Erfassung zu speichern und zu verlinken. Daten müssten erst eine gewisse Verfeinerung durchlaufen, bevor sie geteilt werden könnten. Rohe qualitative Daten beispielsweise könnten auf eine Weise aufgezeichnet werden, die nur der Datenerfasser verstünde. Diese Schwierigkeit beim Erfassen und Teilen von Rohdaten wird erschwert durch die Erkenntnisse aus Work Package One des RECODE-Projekts (<http://recordproject.eu/>). In ähnlicher Weise würden große Datensammlungen vor dem möglichen Teilen die Korrektur statistischer Irrtümer nötig machen. Ebenfalls von Belang war der Kontext, in dem die Daten gesammelt wurden: Sie könnten unter dem Versprechen der Vertraulichkeit erfasst worden sein oder um eine Studie (Aufsatz oder Doktorarbeit) abzuschließen, bei der sichergestellt sein soll, dass die Daten für eine bestimmte Zeit nicht publik gemacht werden. Eine weitere Frage war die Zeitnähe der Da-

ten, es bestehe die Gefahr dass sie zum Zeitpunkt der Veröffentlichung für eine Folgenutzung bereits zu sehr veraltet seien. Diese Schwierigkeit betrifft auch eine breitere Anforderung, auf welche die Herausgeber hinwiesen, nämlich die, dass verlinkte Daten in einer Zeitschrift „verwendbar“ und „replizierbar“ sein sollten. Daten werden unstrukturiert, ohne ausreichende Metadaten und in Formaten gespeichert, in denen sie später nicht auffindbar sind. Diese Bedenken müssen bei der Formulierung von Richtlinien bedacht werden. Daten aus publizierten Artikeln können zwar offene Daten sein, sind aber selten verlinkt oder verlinkbar. Wenn es echtes Data Sharing geben soll, sind Initiativen unerlässlich, welche den Zeitschriften Richtlinien empfehlen, die das Teilen in klar spezifizierter und angemessener Form verlangen. Über unsere Arbeit zum Design von Richtlinien wird an anderer Stelle berichtet; der Artikel wurde für JASIST verfasst und inzwischen begutachtet; das Erscheinungsdatum steht noch nicht fest. Hier reicht es, unsere Herangehensweise kurz zu umreißen.

Die Entwicklung einer Modell-Richtlinie

Unsere ursprüngliche Vermutung, viele der Probleme

des Data Sharing ließen sich im Verlauf der Veröffentlichung durch die klare Richtlinien vonseiten der Verlage angehen, wurde durch die Untersuchung nicht widerlegt. Uns beschäftigte daher vor allem das Ziel, eine Modell-Richtlinie zu identifizieren, die anderen Zeitschriften empfohlen werden könnte. Während wir Informationen über zahlreiche Richtlinien verschiedener Zeitschriften sammelten, sah es eine Zeitlang so aus, als würde sich aus der Analyse eine Modell-Richtlinie ergeben. Dennoch kamen wir allmählich zu der Überzeugung, dass sie sich nicht als Basis für eine Modell-Richtlinie eigneten. Insgesamt ließen die bestehenden Richtlinien tendenziell die Verwirrung, manchmal auch die Widersprüche in den bisherigen Vorgaben von Redaktionen und Herausgebern erkennen. Für ein effektives Verfahren mussten wir uns auf die Ansichten verschiedener am Data Sharing interessierter Gruppen konzentrieren. Als erstes lernten wir daraus die folgende Lektion: Die derzeitige digitale Infrastruktur ist im Fluss; zwischen Herausgebern, Speichern und Systemen bestehen so große Unterschiede, dass sich keine klare Ermunterung zum Teilen von Daten ergibt. Uns war klar, dass:

- Herausgeber das Teilen von Daten, auf denen Arti-



Bis zu
20 %
auf unsere
Bestseller bei
Online-
Bestellungen!

Jetzt bestellen:
www.demco.de

DEMCO®

IHR SPEZIALIST FÜR BIBLIOTHEKEN, MEDIENZENTREN UND SCHULEN

- kel beruhen, unterschiedlich handhaben;
- Anweisungen an Autoren, welche Arten von Daten akzeptierbar sind, wo und wann sie im Verlauf des Publikationsprozesses gespeichert werden soll, generell vage sind;
 - Forscher aller Disziplinen nach eigenen Angaben prinzipiell das Teilen von Daten befürworten, dabei jedoch unüberwindbare Hindernisse empfinden;
 - Forscher glauben, von klaren Richtlinien der Herausgeber und Zeitschriften bezüglich Datenformat und Speicherort profitieren zu können;
 - Herausgeber ebenfalls Hindernisse beim Verlinken und Einbetten von Daten empfinden.

Außerdem kamen wir zu dem Schluss, dass es vermutlich oftmals nicht praktikabel ist, alle Daten einzubeziehen, die in einem Artikel veröffentlichten Ergebnisse untermauern. Datenformate und Dateigröße können je nach angewandter Methodik einer Studie erheblich schwanken. Qualitative Forschung liefert Daten in Form von Dokumenten und Text, beispielsweise Notizen über Ausgrabung und Feldbeobachtung, Niederschriften von Interviews oder Berichte. Qualitative Methoden liefern numerische Daten, die auf Verlaufsbögen notiert oder nur zugänglich sind, wenn auch der entsprechende Computercode zugänglich ist. Aus einer Untersuchung lassen sich unterschiedliche Arten von Daten ableiten; in einen Artikel müssten also zusätzlicher Text, numerische Datensätze und digitale Bilder aufgenommen werden, wodurch die Dateien erheblich vergrößern würden. Im Laufe unserer Untersuchung erkannten wir, dass das unzureichende Wissen, wohin die Daten hochgeladen werden sollten, ein größeres Hindernis darstellte. Viele Akteure waren mit Datenspeichern nicht vertraut, und die übrigen zeigten sich besorgt über deren generelle Infrastruktur. Das bedeutet offensichtlich, dass Daten-Richtlinien einer Zeitschrift etwas darüber aussagen sollten, ob die Daten an einem angegebenen Speicherort mit vertrauenswürdigen Speicherungsrichtlinien gespeichert werden sollten, ob eine dauerhafte URL benutzt und ob ein bestimmter Zitier-Stil angewendet werden soll. Der Zeitpunkt der Freigabe wirft eine interessante Frage auf: Forscher waren weniger darüber besorgt, zu welchem Zeitpunkt im Verlauf der Publikation, sondern zu welchem Zeitpunkt ihrer Untersuchung die Daten offen zugänglich gemacht werden sollten. Oft werden Artikel nicht nur am Ende einer Studie verfasst, sondern auch im Verlauf der wissenschaftlichen Untersuchung. Es kann angemessen sein, die Daten zusammen mit dem Artikel zu veröffentlichen, wobei verschiedene Dinge beachtet werden müs-

sen, beispielsweise die für Doktorarbeiten geltende Voraussetzung, dass eine wissenschaftliche Untersuchung einzigartig ist, oder der sensitive Charakter bestimmter Daten und ethische Überlegungen des Datenschutzes von Personen.

Unserer ursprünglicher Entwurf des JoRD-Projekts umfasste die drei Fragen Wo? Was? und Wann? Das heißt: Wo die Daten gespeichert werden sollten, in welchem Format und zu welchem Zeitpunkt im Publikationsprozess, wobei manchmal auch Sperrfristen für die Freigabe der Daten zum richtigen Zeitpunkt während der laufenden Forschung festgelegt werden können. Die Behandlung sensibler Daten wurde nicht ausdrücklich angesprochen. Die ursprüngliche Richtlinie erwähnte den Datenverweis beiläufig im Rahmen von anderen Anleitungen zu Daten, doch interessierte Akteure brauchen eine klare Aussage über das Zitieren von Daten und Metadaten. In ähnlicher Weise wurde deutlich, dass viele Bedenken von Akteuren über Urheberrechte an Daten ausgeräumt werden müssen, indem Empfehlungen für Metadaten in Verbindung mit Autoren, beispielsweise durch Digitale Object Identifiers, aufgenommen werden. Ein weiterer Modellrahmen für Richtlinien zu Forschungsdaten wurde aus den oben zitierten Einsichten entwickelt. Er soll als eine Art „Richtlinien-Motor“ dienen, aus dem eine Zeitschrift die für sie angemessene Richtlinie entwickeln kann. Uns schwebt ein Verfahren vor, durch welches solche Richtlinien in Kooperation zwischen Forschungsförderern und Forschungseinrichtungen auf der einen und Herausgebern auf der anderen Seite entwickelt werden. Falls dabei Schwierigkeiten auftreten, wäre ein Lösungsverfahren erforderlich, das letztendlich das Recht der Forschungsförderer anerkennt, über die Verwendung der durch die Untersuchung generierten Daten, für die sie (oder vielmehr die Öffentlichkeit) bezahlt haben, zu entscheiden. Wie oben erwähnt, wird die Modell-Richtlinie in einem Artikel beschrieben, der das Peer Review durchlaufen hat und in absehbarer Zeit veröffentlicht wird.

Schlussfolgerungen

Dargelegt wurden die Grundprinzipien zum Data Sharing bei wissenschaftlichen Untersuchungen. Endgültig lässt sich die Frage nicht beantworten. Es fehlen jedoch gegenwärtig die Mittel, das Data Sharing effizient zu machen. Die Autorinnen und Autoren dieses Artikels sind zutiefst überzeugt, dass wissenschaftliche Fachzeitschriften mit Richtlinien über Data Sharing in den Publikationsprozess eingreifen sollten. Durch das JoRD-Projekt konnte der Inhalt von derzeit bestehenden Richtlinien in Fachzeitschriften

gesammelt werden und auf der Grundlage qualitativer Forschung Ideen für den Entwurf von Richtlinien entwickelt werden. Wir präsentieren Problemstellungen, die bei der Entwicklung von Richtlinien zum Data Sharing beachtet werden sollten, damit Forscher zweifelsfrei verstehen, was von ihnen verlangt wird, um nicht nur den Anforderungen der Zeitschrift zu genügen, sondern auch denen der Herausgeber und der gesamten Forschungs-Community. ■

Literatur

- ICSU (International Council for Science (2004) ICSU Report of the CSPR Assessment Panel on Scientific Data and Information. Paris: ICSU.
- Kuipers, T. und van der Hoeven, J. (2009) PARSE: Insight into issues of permanent access to the records of science in Europe. Survey report. Brussels: European Commission.
- McCain, K. (1995) Mandating sharing: journal policies in the natural sciences. *Science Communication* 16, 403-431.
- National Academy of Sciences (2003). Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences. Abgerufen am 3. März 2014 von der URL: <http://www.nap.edu/catalog/10613.html>
- NISO (2013) NISO RP-15-3013, Recommended Practices for Online Supplemental Journal Article Materials. Abgerufen am 14. Januar 2014 von: www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf
- OECD (Organisation for Economic Co-operation and Development) (2007) OECD Principles and Guidelines for Access to Research Data from Public funding. Paris: OECD.
- Piowar, H. und Chapman, W. (2008a) A review of journal policies for sharing research data In: *Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing (ELPUB)*, 25.27. Juni 2008, Toronto Canada. Abgerufen am 3. März 2014 von URL <http://ocs.library.utoronto.ca/index.php/Elpub/2008/paper/view/684>
- Piowar, H. und Chapman, W. (2008b) Identifying data sharing in biomedical literature. *AMIA Annual Symposium Proceedings*, 596-600. Abgerufen am 3. März 2014 von URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PM2655927>
- Piowar, H. und Chapman, W. (2010a) Public sharing of research datasets: a pilot study of associations. *Journal of Informetrics* 4(2) 148-156. Abgerufen am 3. März 2014 von URL <http://www.sciencedirect.com/science/article/pii/S1751157709000881>
- Piowar, H. und Chapman, W. (2010b) Recall and bias of retrieving gene expression microarray datasets through PubMed identifiers. *Journal of Biomedical Discovery and Collaboration* 5, 7-20. Abgerufen am 3. März 2014 von URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2990274>
- Piowar, H. (2010) Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS One* 6:7 07. Abgerufen am 3. März 2014 von URL: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0018657>
- Royal Society (2012) Science as an open enterprise: summary report, June 2012. London: Royal Society. Abgerufen am 3. März 2014 von URL: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE-Summary.pdf
- Shrager, D. et al. (2006) The content of medical journal instructions for authors. *Annals of Emergency Medicine* 48(6), 742-749.
- Smit, E. und Gruttemeier, H. (2011) Are scholarly publications ready for the data era? Suggestions for best practice guidelines and common standards for the integration of data and publications. *New Review of Information Networking* 16(1) 54-70.
- Smit, E. (2011) Abelard and Heloise: why data and publications belong together. *D-Lib Magazine* 17(1-2). Abgerufen am 3. März 2014 von URL: <http://www.dlib.org/dlib/january11/smit/01smit>
- STM (International Association of Scientific, Technical and Medical Publishers) (2007) Brussels Declaration. Abgerufen am 3. März 2014 von URL: <http://www.stm-assoc.org/brussels-declaration/>
- Stodden, V. et al. (2013) Towards reproducible computational research: an empirical analysis of data and code policy adoption by journals. *PLOS One*, 21. Juni 2013. Abgerufen am 14. Januar 2014 von: [doi;10.1371/journal.pone.0067111](https://doi.org/10.1371/journal.pone.0067111)

Hinweis:

Das JoRD-Projekt wurde finanziert von UK JISC (www.jisc.ac.uk).

.....
**Paul Sturges, Marianne Bamkin,
 Jane Anders, Azhar Hussain**

Centre for Research Communication

Nottingham University

Nottingham NG7 2NR, UK.

r.p.sturges@lboro.ac.uk
